FCLib: A Library for Building Data Analysis and Data Discovery Tools

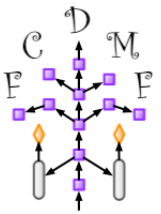
Wendy S. Koegler & W. Philip Kegelmeyer

Sandia National Laboratories, USA

IDA 2005

September 9, 2005

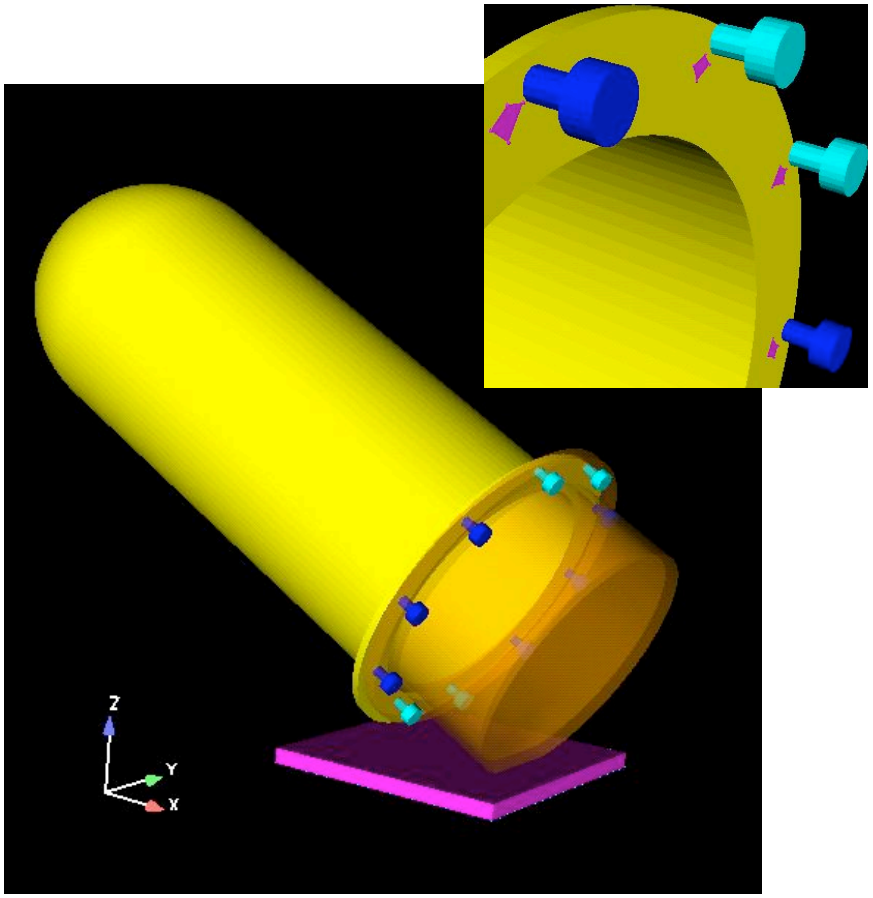
Madrid, Spain



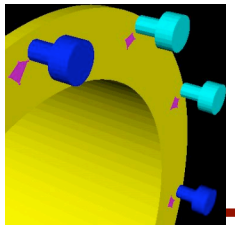
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Example: Evaluating Broken Spotwelds

- How many break & when?
- Original method:
  - Plot force vs. time & manually search for zero crossings.
- New method:
  - Fully automated
  - ‘Close to’ failure measure
  - Used instantaneous model orientation



Sandia National Laboratories

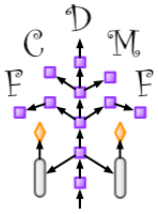


Motivation

- Application Area: Data discovery (DD) in large-scale (terabyte) finite element simulations with focus on feature-based analysis.

- Issues:
  - No longer possible to examine all data
  - Select/Process regions of interest (ROIs)
  - Difficult to move data between tools
  - DD is a process

- Solution:
  - Analysts need more automation
  - … but they still need flexibility



Feature Characterization Library (FCLib)

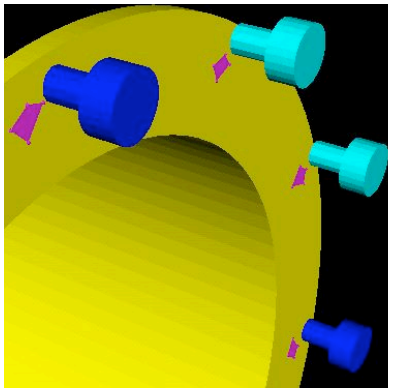
Toolkit of data analysis building blocks (C library)

Goals:

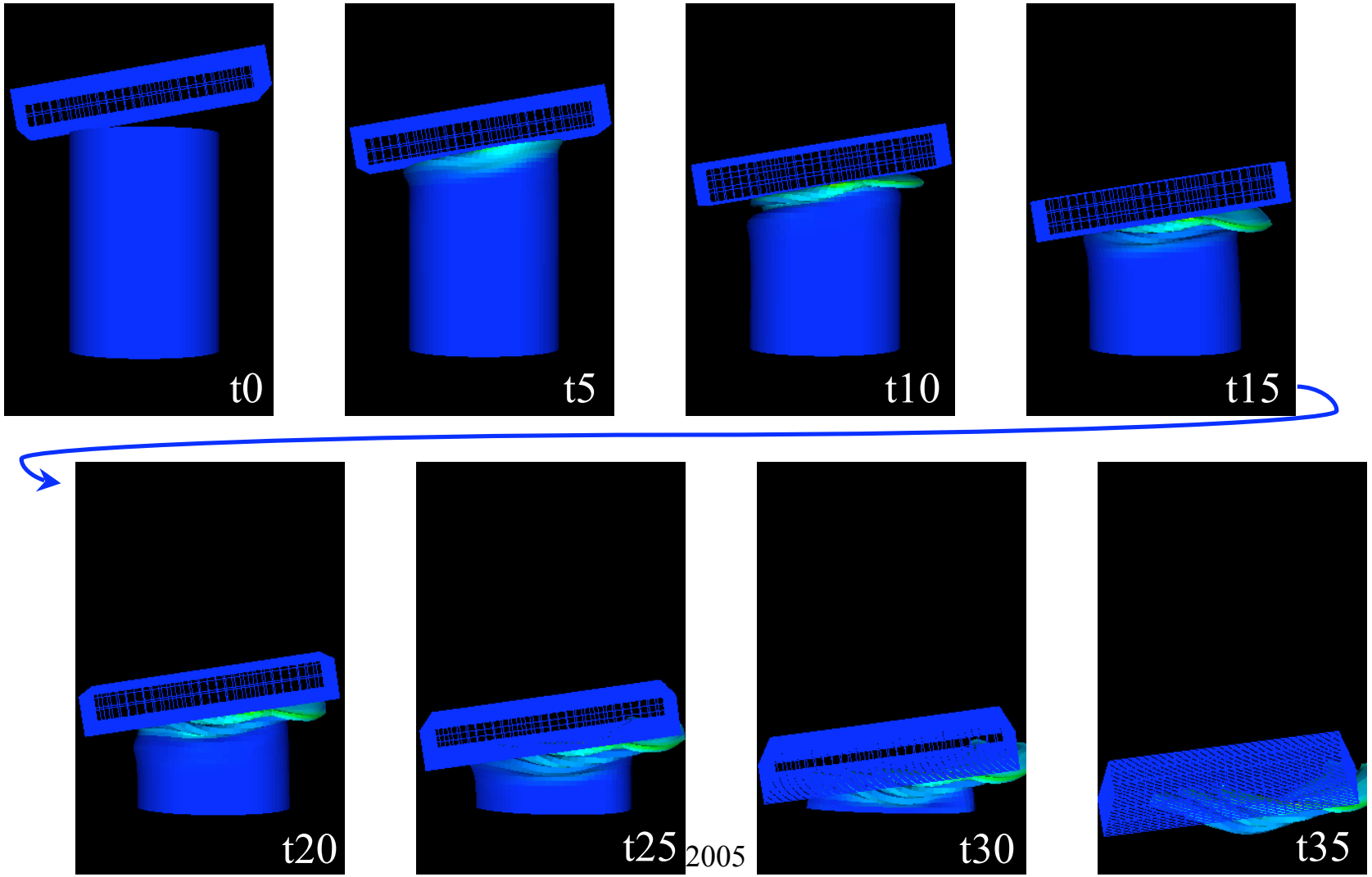
- **Provide variety of simple building blocks that can be composed into complex analyses**
- **Support feature-based analysis (feature = region of interest)**
- **Minimize low-level processing (automation)**
- **Generality - applicable to a variety of science domains**
- **Provide an 'elegant' interface - simple but not constrained**

Presentation Outline

- Motivation
- Goals
- Toolkit Design
  - Data Model
  - Analysis Building Blocks
- Data Analysis Examples
- Summary & Conclusions

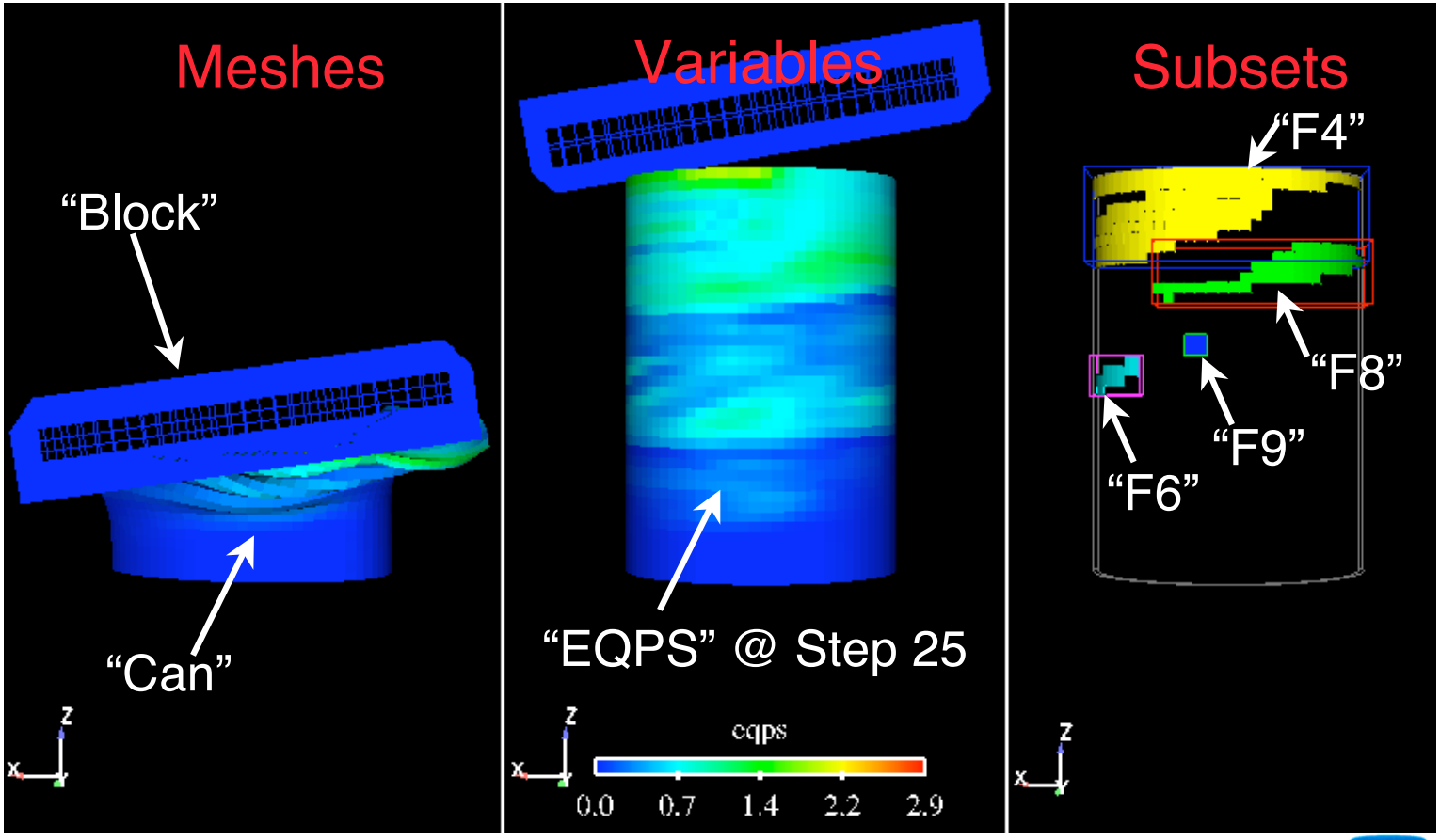


Example Data: Can Crush



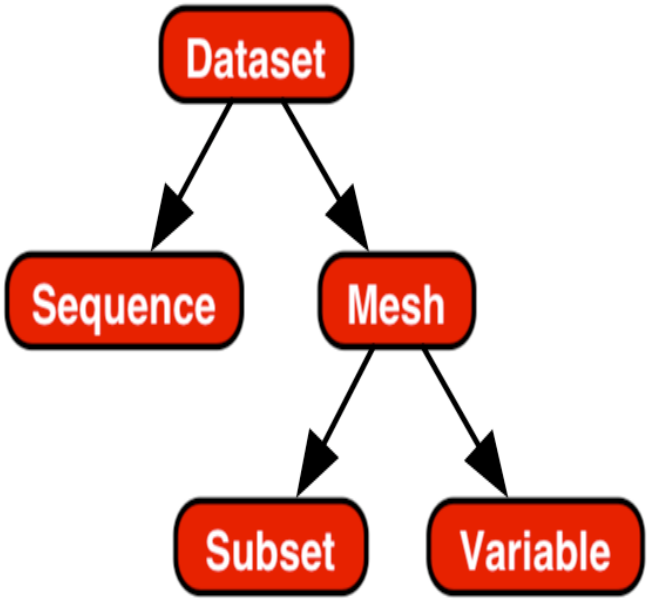
Dataset = “Can Crush”

Sequence = “Time”; coordinate @Step 25 = 0.0025



Sandia National Laboratories

Key Features of the Data Model

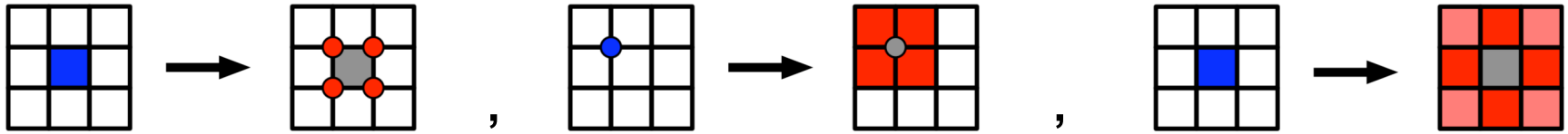


- Minimal set of abstract data objects
  - Types are deeper in the interface
- Subsets are full-fledged data objects
- Mesh ownership of data
- Time-varying data available per variable instead of per step

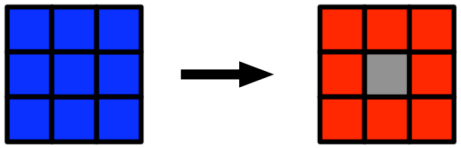
- Evaluated data model libraries:
  - Parallel Mesh Object (PMO) by TeraScale
  - Visualization Toolkit (VTK) by Kitware
  - Data Object Library (DOL) by Sandia

Analysis Building Blocks I

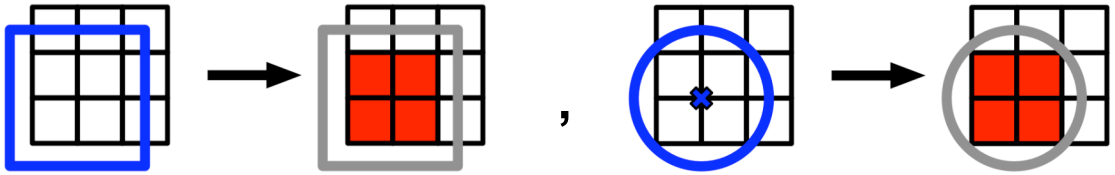
- Mesh Topology
  - Get mesh entity children/parents/neighbors



  - Segment into separate connected components
  - Get skin

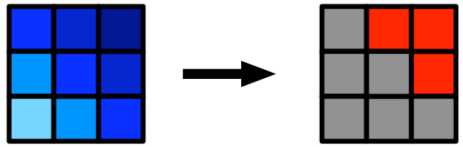


- Mesh Geometry (Spatial)
  - Edge lengths/surface areas/region volumes
  - Get mesh entities within bounding box or sphere

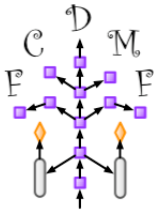


Analysis Building Blocks II

- Variable
  - Variable Math
  - Statistics: Min/Max/Mean/StandardDeviation/Sum
  - Decompose vectors into normal & tangent components
  - Threshold: find subset

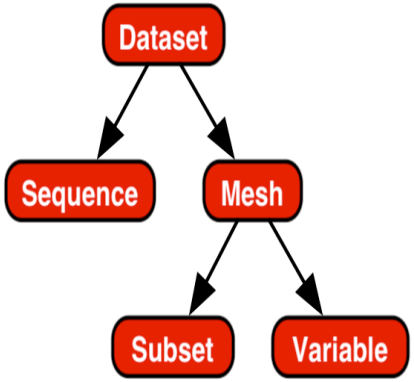


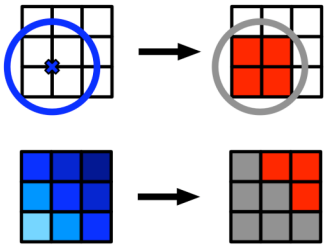
- Subset
  - Set operations: AND, OR or XOR to get new subset
- Time
  - Feature tracking - track ROIs over time



Presentation Outline

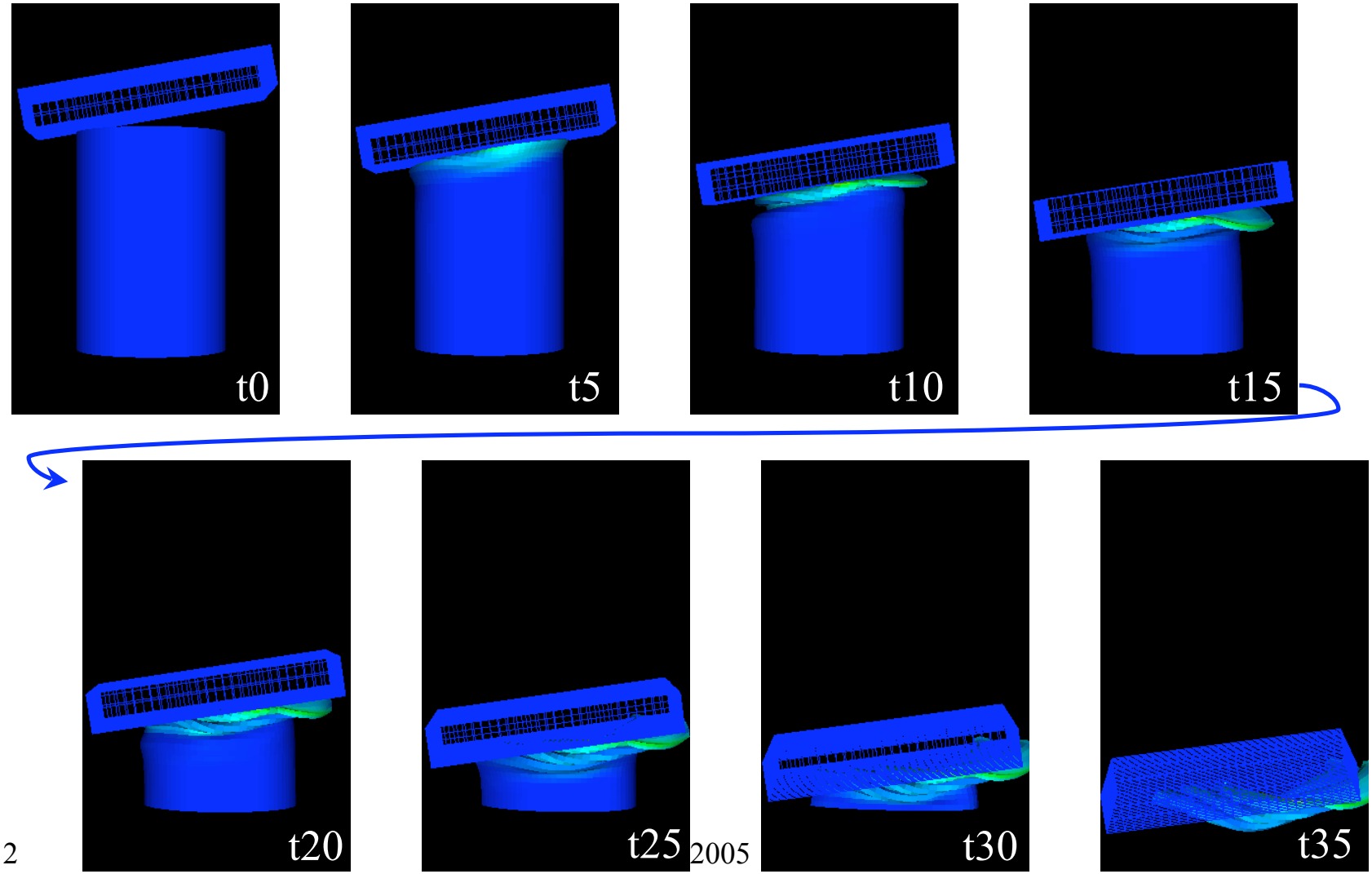
- Motivation
- Goals
- Toolkit Design
  - Data Model
  - Analysis Building Blocks
- **Data Analysis Examples**
- **Summary & Conclusions**

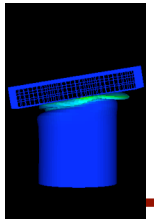




Sandia National Laboratories

Example 1: What's Going on in the Ridges?



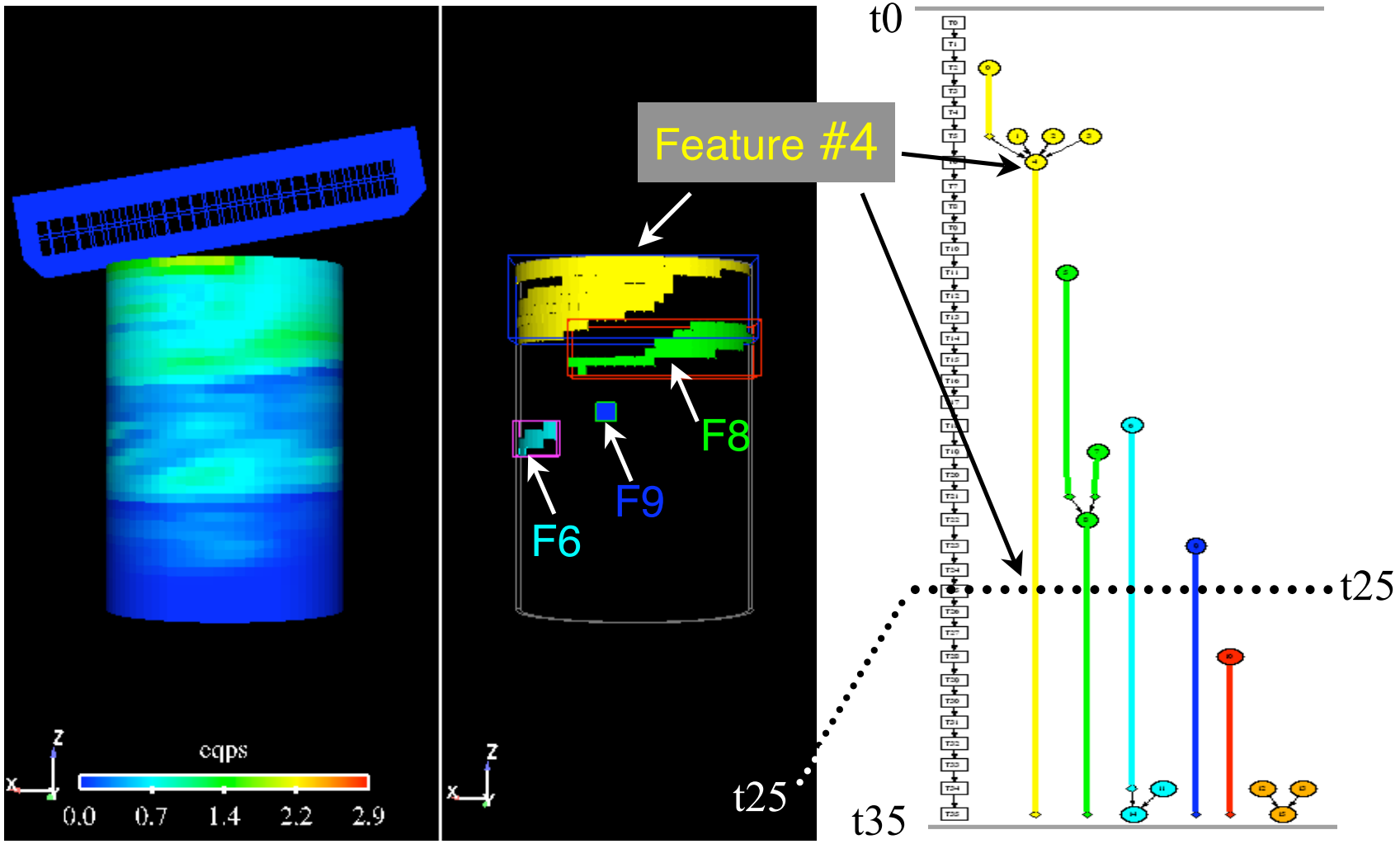


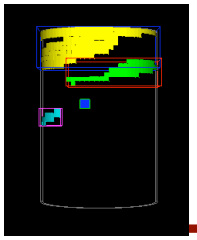
Find & Track Features

```
fc_loadDataset("cancrush.saf", &dataset);
fc_getMeshByName(datatset, "can", &mesh);
fc_getSeqVariable(mesh, "eqps", &numStep, &seqVar);

fc_createFeatureGroup(&featureGroup);
for (i = 0; i < numStep; i++)
  {
   fc_threshold(seqVar[i], ">=", 1.0, &subset);
   fc_segment(subset, 0, &numROI, &ROIs);
   fc_trackStep(i, numROI, ROIs, featureGroup);
  }
fc_writeFeatureGraph(featureGroup, "graph.dot");
```

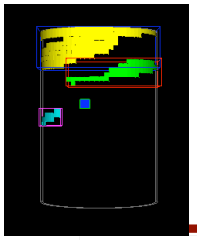
Reading the Feature Graph



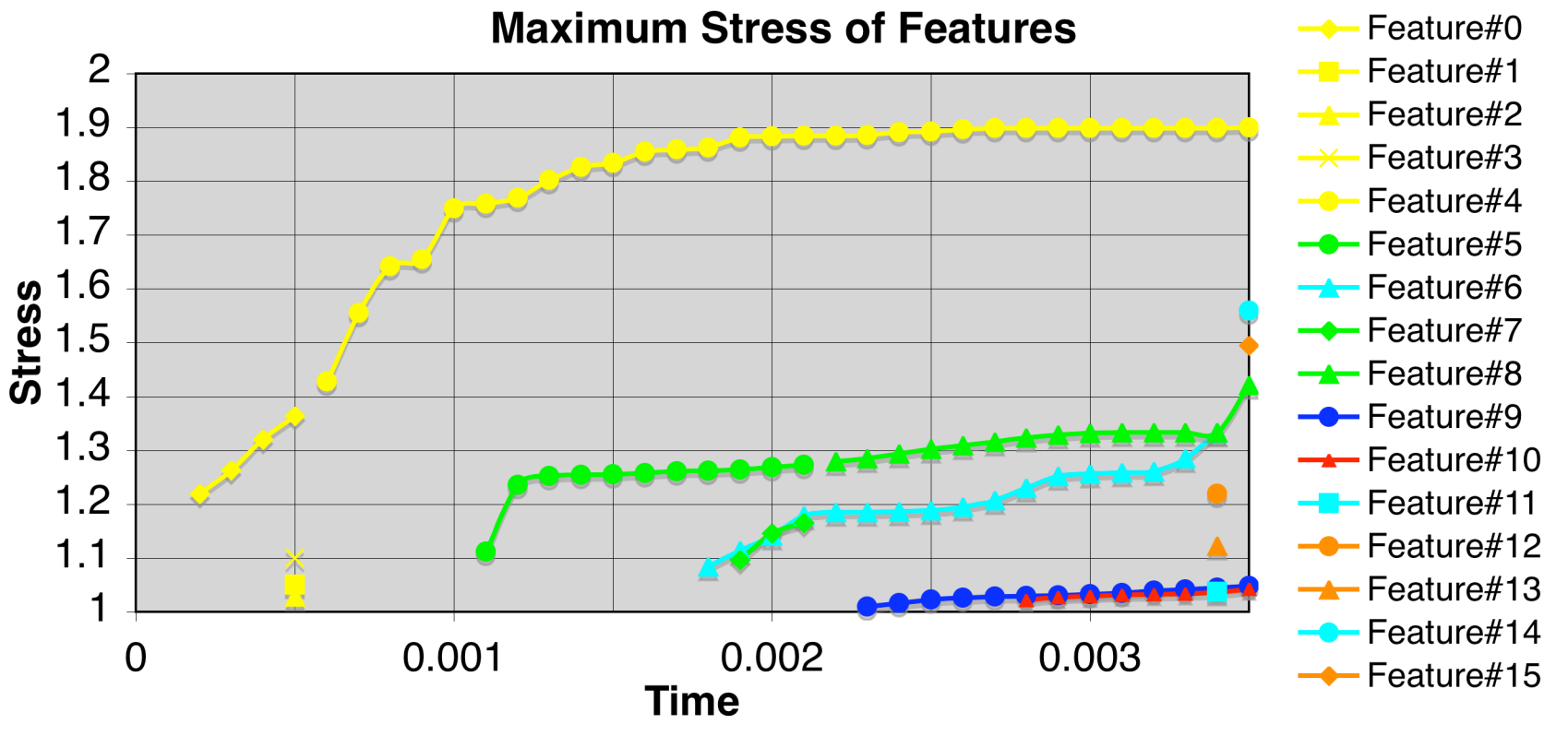


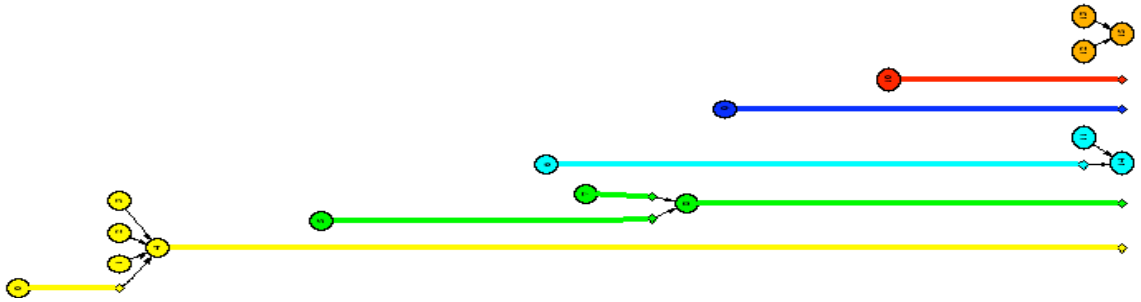
Analyze the Features

```
fc_featureGroup_getNumFeature(featureGroup,
                                  &numFeature);
for (i = 0; i < numFeature; i++)
  {
   fc_getFeatureROIs(featureGroup, i, &numROI,
                       &stepIDs, &ROIs);
   for (j = 0; j < numROI; j++)
    {
     fc_getVariableSubsetMinMax(seqVar[stepIDs[j]],
                  ROIs[i], &mins[i][j], &maxs[i][j]);
     // print stats
    }
  }
```



Interpret Results & Repeat



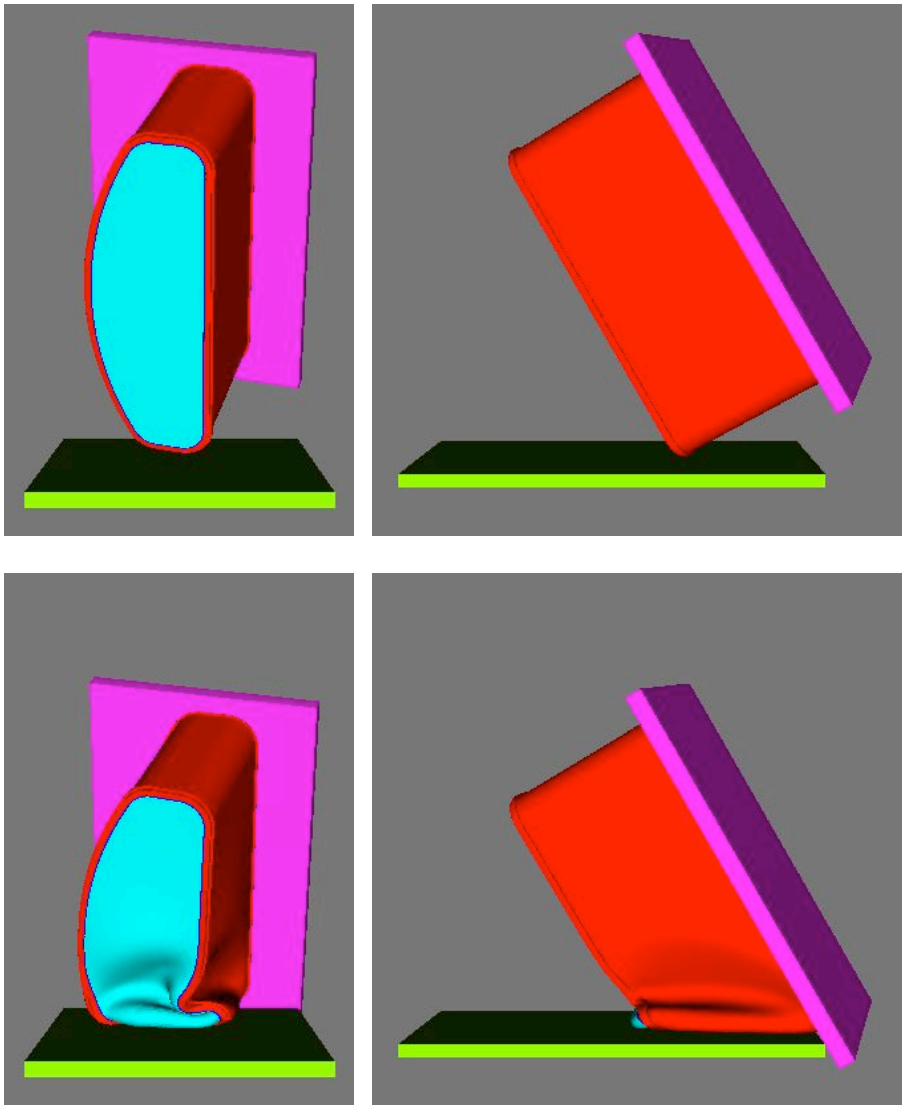


Sandia National Laboratories

Example 2: Yet Another Can Crush Study

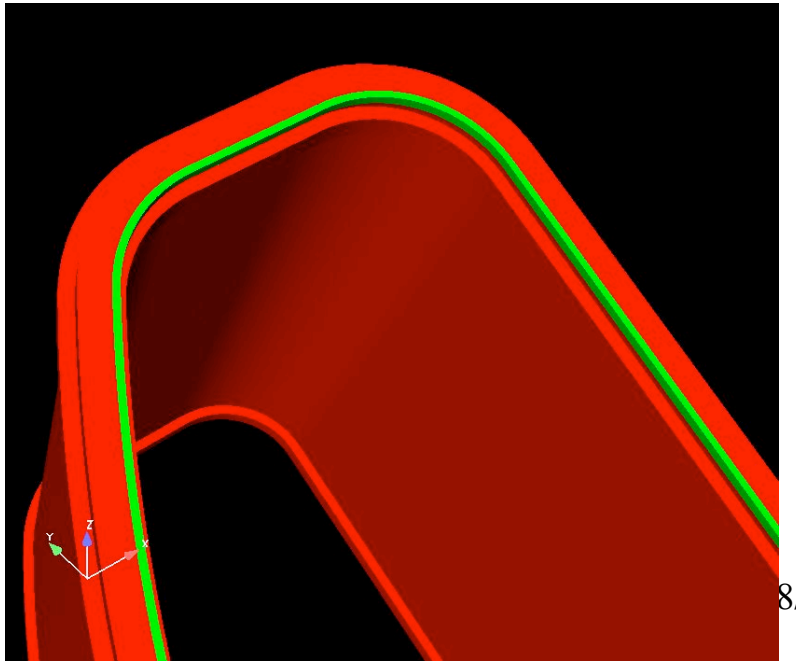
- Complex assembly
- Need to assess
  - simulations vs. experiments
  - effect of parameters
- Want to predict behavior in even more complex systems

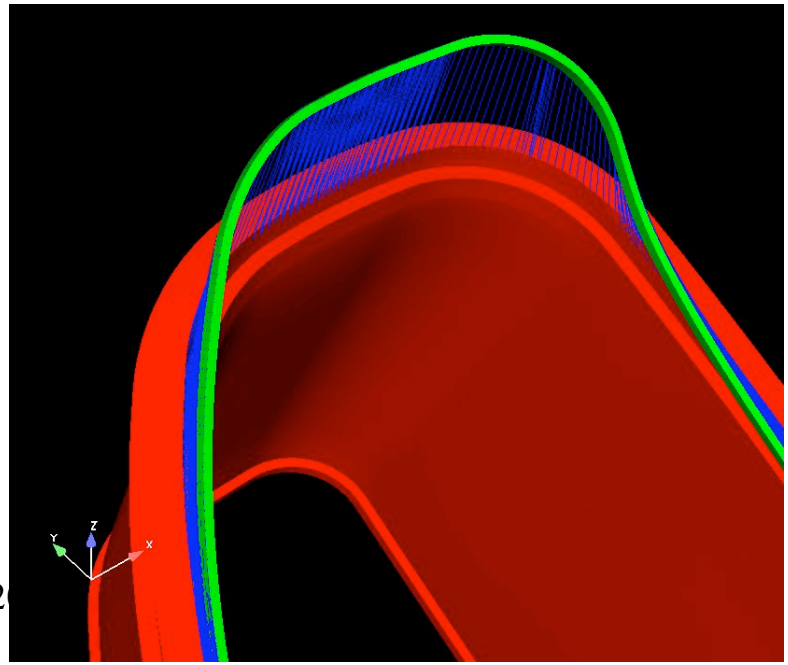
- But what to measure?!
- FCLib = prototyping environment

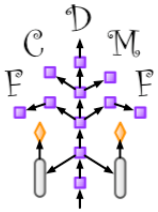


Analyses In Progress: Dents, Tears & Gaps

- Dents (deformed regions) - Subtract translation from general displacement
- Tears (regions of failed elements) - Use topology to determine type and if breaks the mesh.
- Gaps (adjacent meshes separate) - “Gap lines” mesh.

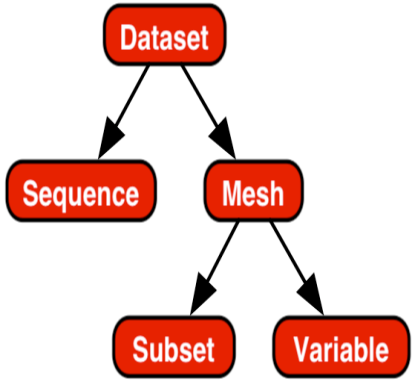






Summary & Conclusions

- Presented and discussed a data analysis toolkit

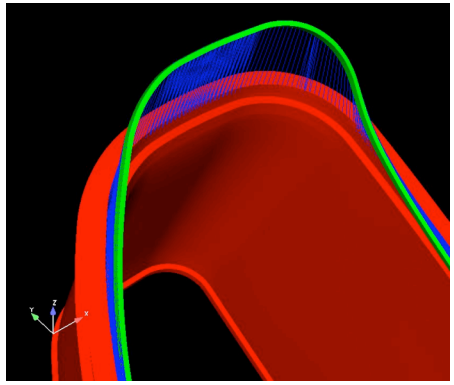


Goals:

- Provide variety of simple building blocks than can be composed into complex analyses
- Support feature-based analysis (feature = region of interest)
- Minimize low-level processing (automation)
- Generality - applicable to a variety of science domains
- Provide an 'elegant' interface - simple but not constrained

Future Plans

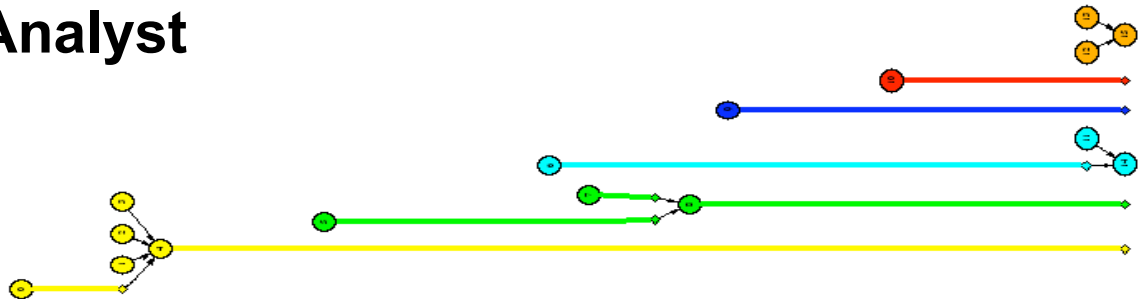
- More building blocks & analyses:
  - Dent, gap, & tear
  - Time-based & event
- GUI for analysis composition
- Better integration with viz packages



- Use FCLib for research:
  - "Smart" dataset comparisons using characterized features

Acknowledgements

- Ann Gentile, FCDMF Developer
- Jay Dike, Analyst



Contact: Wendy (Koegler) Doyle <wkoegle@sandia.gov>

http://www.ca.sandia.gov/ASCI/sdm → Data Discovery → FCDMF

This research was supported by the DOE Advanced Strategic Computing Initative (ASCI) Data and Visualization Sciences (DVS) Data Discovery Program.

Sandia National Laboratories